Identification of a novel class of annexin genes

Vahid Khalaj[1], Lyndsay Smith[1], Jayne Brookman, Danny Tuckwell*

F2G Ltd., P.O. Box 1, Lankro Way, Eccles, Manchester M30 0BH, UK

Received 17 December 2003; revised 16 February 2004; accepted 16 February 2004

First published online 27 February 2004

Edited by Takashi Gojobori

Abstract The annexins are a family of calcium- and phospholipid-binding proteins that have been widely studied in animals. Investigation of annexins in the fungus *Aspergillus fumigatus* identified a novel annexin-like gene (ANXC4) as well as two conventional annexins (ANXC3.1 and ANXC3.2). The genes were initially identified by bioinformatics, and sequences were then determined experimentally. Reverse transcription polymerase chain reaction indicated that all three genes were expressed. ANXC4 lacked calcium-binding consensus sequences and had a 553 residue N-terminal tail. However, bioinformatics indicated that ANXC4 is an annexin and homologues were identified in other filamentous fungi. ANXC4 therefore represents a new grouping within the annexin family.

Key words: Annexin; Duplication; Calcium; Pathogen; *Aspergillus fumigatus*

1. Introduction

The annexins are a family of Ca^{2+}- and phospholipid-binding proteins found throughout the eukaryotes [1–3]. Each annexin can be divided into a conserved ~35 kDa core domain, which interacts with membrane phospholipids in a Ca^{2+}-dependent manner, and a variable N-terminal tail, which interacts with other annexins or different proteins [1]. The core is made up of four ~70 amino acid repeats which contain the consensus endonexin sequence GxGT...D/E [4]. Sequences specifying 'type II' and 'type III' Ca^{2+}-binding sites as well as a phosphatidylserine-binding sequence have been identified [5,6].

A wealth of observations including co-crystallisation of annexins with phospholipids [7,8], reports of membrane binding and aggregation by annexins (e.g. [9,10]) and the Ca^{2+}-dependent association of annexins with endosomal membranes [11] have led to annexins being implicated in a diverse range of functions including exocytosis and membrane fusion, as well as Ca^{2+} channel regulation [1,12,13]. At least 13 annexins have been identified in humans, with different functions described or proposed for different annexins [14].

A fungal annexin has previously been identified in *Neurospora crassa* by bioinformatics [2], and we have identified an annexin in *Aspergillus niger* (accession number AY033935_1; V. Khalaj, G. Robson, J. Brookman, manuscript in preparation). Here we describe the identification and sequencing of three novel annexin genes (ANXC3.1, ANXC3.2 and ANXC4) in the human pathogenic fungus *Aspergillus fumigatus* and show that ANXC4 is unlike any previously reported annexin.

2. Materials and methods

2.1. Database searches

The *A. fumigatus* genome (The Institute for Genome Research; www.tigr.org) was searched with the ANXA protein sequence (AY033935_1) using tblastn. The corresponding DNA contig sequences were downloaded from TIGR and gene predictions carried out using Genscan and Genefinder (Genscan: genes.mit.edu/GENSCAN.html; settings: organism = vertebrate; suboptimal exon cutoff = 1.00: Genefinder, genomic.sanger.ac.uk/gf/gf.html; FGENESH (with donor GC); settings; organism = yeast). Predictions were visualised in Artemis (http://www.sanger.ac.uk/Software/Artemis/ [15]) and modified by alignment with known annexins.

Database searches to identify annexins in other fungi were carried out at the following web sites: *Candida albicans*, www-sequence.stanford.edu/group/candida/search.html (Stanford Genome Technology Center – sequencing was accomplished with the support of NIDR and Burroughs Wellcome Fund); *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and *Encephalitozoon cuniculi*, www.ncbi.nlm.nih.gov; *N. crassa*, www-genome.wi.mit.edu/annotation/fungi/neurospora [16]; EST database, www.blast.genome.ad.jp and cogeme.ex.ac.uk. Searches of the *Aspergillus nidulans* genome along with some *N. crassa* and *Magnaporthe grisea* searches were carried out using downloaded data running on a local machine (Aspergillus Sequencing Project, Center for Genome Research, http://www.broad.mit.edu; Magnaporthe Sequencing Project, Ralph Dean, Fungal Genomics Laboratory at North Carolina State University, http://www.fungalgenomics.ncsu.edu, and The Center for Genome Research, http://www. broad.mit.edu). Searches of ANXC4 against the non-redundant GenBank database (16/7/03 release) and PFAM analysis were carried out locally.

2.2. Preparation of DNA and RNA

A. fumigatus clinical isolate AF293 (the kind gift of Dr David Denning, clinical isolate culture collection, Hope Hospital, Salford, UK) was cultured on Sabouraud dextrose agar (Oxoid). A standard inoculum of 1×10^6 spores/ml (final concentration) was added to 50 ml Sabouraud dextrose broth and grown at 37°C with agitation at 225 rpm. Biomass was harvested in mid-exponential phase by centrifugation (10 min, 3000 rpm). DNA isolation was performed essentially according to [17]. RNA was prepared using the FastRNA kit (QBIOgene) and treated with RQ1 RNase-free DNase (Promega). General recombinant techniques were carried out as described in [18].

2.3. Polymerase chain reaction (PCR) amplification of annexins

gDNA and cDNA were amplified by PCR and reverse transcription (RT) PCR respectively using primers that generated overlapping fragments thereby covering the entirety of each gene. cDNA sequence was generated from RACE-ready cDNA (RACE: rapid amplification of cDNA ends), synthesised using the Invitrogen GeneRacer Kit. Primers were designed to produce fragments covering the length of the predicted genes and untranslated regions (UTRs). For instance, ANXC3.1-specific primers LCS1 (5′-TCAGCAAGTACCAGAAG-

*Corresponding author. Fax: (44)-161-785 1273.
E-mail address: dannytuckwell@f2g.com (D. Tuckwell).

[1] These authors contributed equally to this work.

0014-5793/04/$30.00
doi:10.1016/S0014-5793(04)00186-3

TC-3′) and LCS2 (5′-GATCCTTCGCCTTAGACA-3′) amplified a fragment of 354 bp from both genomic and cDNA; ANXC3.2-specific primers LCS35 (5′-TCCGACTATGTCATACAACC-3′) and LCS4 (5′-CTTCGTCGACGTTGTTGAC-3′), spanning a putative intron, amplified a fragment of 1042 bp from gDNA but a fragment of 968 bp from cDNA; and ANXC4-specific primers LCS5 (5′-CATGAAGGCAGAACTCAAG-3′) and LCS6 (5′-TGCAACAGCAGAGCATCG-3′), also spanning a putative intron, amplified a 415 bp fragment from gDNA and a 351 bp fragment from cDNA. DNA fragments were cloned into pGEM-TEasy (Promega) and used to transform *Escherichia coli* TOP 10 cells (Invitrogen).

5′ and 3′ ends of sequences and corresponding UTRs were determined by RACE, using the Invitrogen GeneRacer kit with the following primers: ANXC3.1 5′, 5′-CTGAGGATACGCTCCCTGAGGCGGATAC-3′; ANXC3.1 3′, 5′-GCCACCTGGAAGATGCGCTACTGTCGAT-3′; ANXC3.2 5′, 5′-GGCATGGCGGGTGGGCCTTGGAACTG-3′; ANXC3.2 3′, 5′-GTCAAGCGCGCGTACCACCATCGGTAC-3′; ANXC4 5′, 5′-GGCACCAGTCATCTCAGGGTAGGTCGACA-3′; ANXC4 3′, 5′-TCGTCGGCGCTATGGGGAACGGCTTGA-3′.

Automated DNA sequencing was carried out by MWG Biotech (Milton Keynes, UK).

A

CGAGACAACCATTAACAAGGGTTCCCATGTTTCATGCATGTGCCAAAGGGTCGGTGGGCTGATCGGGAAATTTGGCTTGGCATACGGCCG
TGCACTGTGGCTTGATAGGTTCCTCATTGGATGAAGCGAGTAAGGACCCTCTTCACTATTGGTCTAAGTCAATTTGTGTGTCTCTAGTGG
CTGTGCATTTCGTAATTCAGCTGGAGATGATGCCAATCAGTGGGACAAATATTTAGCCAAATAGAGGGGGTTAAACTCAACCGATTGCGG
CTGTGTTGACAAGGGCTAAAATACGAGTTGAGCACAGGCTAGATAACTTTAAAATATTATCTGGTGAGGGTGAGGGGGATATAAGGCTGG
CGTGTGCTCCCTTTGTTAAGTCTCGATTGTTGTTACCTTTCCCCTGTCACATTTCTGTC**ATGTCGCACCCCCAGTATACCTACGGCGGAC**
CCCCGCAACCCCCTTATAACAGCGCTGGGTATGCTCAGCCGGGCGCCCATCAACAGCAGCCGGCCCCGTTTTATCCACCACAACAGCCTG
GTGGATATCCGCCCCAAGGGGCTCCCCCGCAGCAATTTCCCCCCTCTGGTCAGTACCAACAGCCATATGGTGCGCCGCCGCCGCAACAAC
AATGGCCGGGCCAACCACCGCAACAATATCCTCAGGGAGGCCATCCACCACCGCAGCAGTATCCGCCTCAGGGAGCGTATCCTCAGAATC
AATACCCTCCACAGGGCCAATATCCTCCCCATGGTCAGTATCCTCCGCAGGGAGGTTACCCGCCGCAAGGCCAACCCGGCTATCCACCCC
CTCAGGCGCCTCAACCAATGCCCACTCCGCCATCCCTAGGCTATGACCCCAACCAACGCGCCCCCGGCGACGCAACCCGCGAAGCCGAAG
CCCTCCGCAAAGCAATGAAAGGCTTCGGCACAGACGAAAAAGCCCTGATCAATATCCTGACCAAGCCCGACCCCCTGCAAATGGCACTCA
TCAGACACACCTACACCGATCGCATCGGACGCAATCTCGAAAAAGACCTCAAGTCCGAGCTCTCGGGGGACCTGGAAGACGTCCTCCTTT
CTTTAGCCAGAGGCCCACTTGAGCAGGACGTCGCTACCCTGCGCGGTGCCATGTCAGGCCTAGGAACGAACGAAAACCTCCTCACGGATG
TGCTGGTCGGCCGGTCGAATGCCGACCTGCGCGCAATCAAGTACGCGTATGTCAGCAAGTACCAGAAGTCGCTGGTGGAGGATATCAAGA
GCGATTTGTCCGGCAAGACCGAGCAGTTCTTTGTCATGTTGCTGAACGCTACGCGGCCGGAGCCGGGGACGTATTTCGATGCGCAGTCGG
TTGATGCTGATGTTCGCGAGATCCACCAGGCGACGCAGGCGCGCAAGGGGACAGATGAGATTGGTGTCTTTGCCGTACTTTTGGGTGCCT
CGGATGCGAAGCTCGTCGCAATCGCGCAGGCTTTTGAGGCAAAGTATCATATTAGCCTCGAGAAGGTGATCAAGGATGAGTTCTCGGGCC
ACCTGGAAGATGCGCTACTGTCGATGCTGTCTAAGGCGAAGGATCCTGTTGGTCATGATGTGGAGAAGCTCCTAAAGTGTTTGCCTCCTG
TTGAGAACGGGAAAGCAGATGTGAAGCGGTTGATCTACTGGGTTGTTCATCTGCATTGGAACCCACCCCATTTTGCAGCGGTCAAGGCGC
GGCTGAAGCAGAAGCTGGGCCATGATGTTGGCAACCACTTCAGAGAAGCCATTGCCCCAGGAGACTTCCTTACTGCTATGCGCCAGGTGT
GGAATTCGGCCTAGACGATTATTATGAATAGGGTTTGGACAAGTCCGGTATTGCTATTCTTAATGTTTATTCTTCTTCTGCAGACCCTCG
AGATTGTGGACGTGGGATTCAGTTTATCGATACTATGGGAGGAATTGTATTTTCAGAAACCAGTTTTGCATTACTGACAGTAGCTTTATA
CGCGCTACTCCTTGGATGATTTTTAGGACTTGGATTTTGTAGGAAGAAAGTACATTACTGTCTCTATTATTCCATGACAGCGCTTGAATT
GATCTTTGACTTAACTCTTATCTGAATCATGGCAGTTATTTTATGAGCTCAGGTCACCTCACCTTTTGTGTAACTACGGGTTCAGTCAAC
CTTTACTAATGTGGGTCAGCCAATCTTTATAAAAATGGACATATTCTGTAACACCTTTGAGTTTTACAGGCCTTTCACCATCTAAAAGAT
AGAAAGATAACTGAACTCCTAAAGATAGTGACAGTTATAAACACATTTGAAGTAATTAAAACCCTTTGGCCCTTAGATAGTATTAAAGAT
CGTAAGGTATCGTATTAGC

B

AGCCGACATATGACGACGATGTACCCCGTAGACAGGAACTAATTTCATCATATAAGTTGACCATTGTCCCTGATTCTCGCTACTTTTGAT
TATATCCTCTCCTTGTCAAATCTTTAGGATCCTTGATATACAGTCCACAAACTGAAAGGAATCCGACT**ATGTCATACAACCAATATCCCC**
CACCAAATCGTATGCCCCCCCCGCATGCCTATTAGAATTCTGCTGCCTATTATAACTGGATAGACTGACTGACTCCTCTGCAG**AACCCCC**
CTACCCTCAGTACGGGGCTCCTCAAGGCTTTTCACAACCGCCTTACCCCCAACAATCGGGTTACGGTGGTTTTCCCCTTCCTCAGGGCCA
CTATAATCGACCACCCCCGCCTGCCCCGTCGCCAGGCGGATATCCTCCGGCTCCCGGACCAGGATATCCCCATTCGCCGCCGCCTCAGCC
TCCATATGGCGCTCCTTCGCAACATCCGTACCCTCCACAAGGCGGTCCTGGATACCCTCCACCTGCGGGCGGGTATCCCCAACCCGGACC
CTATGGCGCCCCACCGGTAGGAGGGAGCGGCTATCCCACACCACCACCGCAACAGTTCCAAGGCCCACCCGCCATGCCTTCTCTGGGTTA
TGTTCCTGGCCAAATGGCACCGGGCGACTTCCGCCGTGAAGCAGACCTCCTCCGCAAAGCGATGAAAGGGTTCGGTACTGACGAGAAGAT
GCTCATCCAAGTACTGTCTAAACTCGACCCCCTGCAAATGGCCGCCGTTCGGTCCACCTACACCAACCATCACCACCGGGACCTCTACAA
GGACGTCAAGTCCGAAACAAGCAGCTACTTCCGCCAGGGCCTGCTCGCCATCATCGACGGACCGCTCCTTCACGACGTGCAGTCCCTCCG
CGAAGCAGTCCAAGGCCTCGGCACAAAGGAATGGCTCTTGAACGATGTCGTCCTGGGCCGCTCAAACGCCGATCTCAACGCCATCAAGGC
CGCCTACGAGCACACATTCCACCGATCCCTCCAAAAGGACGTCGAAGCCGACCTCTCCTTCAAAACCCGCTCCCTCTTCTCCCTCGTCCT
CCGCGCAGAACGCCACGAACCCTCTTACCCGATCAACCCCCAGCTCATCGAGCAGGAAGCCCGCGCCATCCACGCCGCAACCAGCGGCCG
CGTCGTCAACAACGTCGACGAAGTCTGCGGCATCTTCGCCCGCGCGTCGGATCCGGAGCTCCGCGCCATCAGCCAGGCTTTCGGCGCTCG
GTACAACTCGTCGCTGGAGAGCCACATCGAGAAGGAGTTCTCCGGGCATATGAAAGACGCCCTGCTGCATATGCTCCGCACCGCGCTGGA
CCCGGCCATGCGGGACGCGGACCTGCTCGAGGACTGCATGAAGGGGATGGGCACCAAGGACGAGAAACTGGTTACTCGGGTCGTGCGCTT
ACACTGGAACAGACAGCATCTCGACCAAGTCAAGCGCGCGTACCACCATCGGTACAAGAGGGATCTCATCGCCAGAGTCCGCGGAGAAAC
TAGCGGGGATTACCAGAAGTTAATGGTTGCGCTGCTCGAGTGATTATGATTTCTTCTTCGATTCTTTCCATGTCGTTCTAACCTTTATTA
CTGGGTGGTTTCTTGCTTTCGTTGCCATACCATGCTGGCTTACTTCTTTGCTGTGATACCACGATGATACGTTTTATGGCTTGATGTGTT
TTGGCGGGTACAGGCCCGCCATGATTGAAACACAATATGATTGAGCAACTGGGTCGTTGCTCAAATGGAGATAGTCCTAGCTAGAGAAAA
CCTATGAAATACAACTCGATGAAATGCATTTGCTCGCATAAATCTCCTGTTTCTGTCGCATTATGTTCACATTATACATCACACATCAAT
CGTAAGTCGTATCCCATGCAATACTC

Fig. 1. Genomic and cDNA sequences for *A. fumigatus* annexins. A: ANXC3.1. B: ANXC3.2. C: ANXC4. Genomic DNA sequence was determined experimentally. cDNA sequence was determined by cloning and sequencing PCR products using RACE-ready cDNA as a template. 5′- and 3′-UTRs and start and stop codons were determined by RACE. The experimentally determined gDNA sequence of ANXC4 started at base 494 of the sequence given in C. Mature mRNAs are shaded and coding regions are given in bold. TATA and CAAT boxes and polyadenylation sites were identified using programs in WebGene (http://www.itba.mi.cnr.it/webgene/) and are shown underlined where they fall within the experimentally determined region.

2.4. Southern analysis

Southern blotting was carried out as described in [17]. *A. fumigatus* genomic DNA was digested with restriction enzymes, run out on a 1.5% agarose gel, transferred to nylon membranes and probed with the following *A. fumigatus* annexin fragments: ANXC3.1, 1.2 kb, corresponding to bases 407–1575 in Fig. 1A; ANXC3.2, 1.8 kb, corresponding to bases 1–1788 in Fig. 1B; ANXC4, 1.8 kb, corresponding to bases 1691–3488 in Fig. 1C. Probes were digoxigenin-labelled using the DIG-High Prime DNA Labeling kit (Roche). Southern blots were washed at moderate (2×SSC washes) or low stringency (3×SSC washes).

2.5. Phylogenetic analyses

Phylogenetic analysis of sequences was carried out using PHYLIP [19], essentially as described in [20], with the exception that distance matrices were generated using PROTDIST with the Jones–Taylor–Thornton model and trees were inferred from distance matrices of bootstrap replicate data using NEIGHBOUR. The multiple alignment used for the analyses was that given in Fig. 2A with *A. nidulans* sequences added, without the variable N-terminal tail, and with gaps removed. Trees were viewed using TREEVIEW [21].

3. Results and discussion

3.1. Identification of annexin genes in A. fumigatus

The *A. fumigatus* genome (www.tigr.org) was searched with the *A. niger* annexin protein sequence (GenBank AY033935_1) resulting in three matches, with *E* values of 1.5e-88, 4.9e-64 and 1.8e-11. Following analyses of these regions of the genome, three annexin genes were identified. A unifying annexin nomenclature has been proposed by Morgan and Fernandez covering all eukaryotes ([22,23], www24.brinkster.com/annexins/seq/sequences.asp). Taking into account the relationships between the three new annexins (see below), the new proteins were given the next available numbers in the ANXC grouping (fungi and mycetozoa), and are henceforth referred to as ANXC3.1, ANXC3.2 and ANXC4 (ANXC1 and ANXC2 correspond to *Dictyostelium discoideum* annexins [23,24]).

gDNA and message sequences were experimentally determined for all three genes: gDNA constructs were generated by PCR and sequenced, cDNAs were cloned and sequenced, and the 5′ and 3′ ends of the genes were determined by RACE. Comparison of the gDNA and message sequences enabled the identification of introns. Gene sequences are given in Fig. 1 and the deduced protein sequences in Fig. 2. Southern blotting was carried out using fragments corresponding to each of the three genes as probes (Fig. 3). This demonstrated that only a single copy of each gene was present in the *A. fumiga-*

C

GCCAGCGCAAGATGAGTGCGATCTCTTCGCAATGATTGAAATACTGGAAAACAGGAGGAGAGATATAGATGGTTGTCTCTTTTCTTGACT
TTCAACGCAGAACATGCTTATCGGTTTGTAACCAGCCCCGCGAATTCCAGCGCGGGGGCGAAACTACCTCGACCTGTTGTGTGTGTGTGA
TTCCTCCCATTCTTAGCTTCAACCTGTTCAGCTGGACTTTGGCGCGGTCGCTGATGATCACGGTAGAATTCAGCCTCGGGCCAGTAACAA
AAAGAAAGTGGTTCGCTCGAGCCGCTTATGGGTTCTGAGCCTTTGACTGATCTTTCCTCTGTTCTCTCTTGCTCACATTTCTGATCTGCT
TACCCTCCATTCGCTGTTTTGTAGACCTTGACTTTTTCGATTGTTTCGCCTGTATTTCCCCCTCATTGCTTCATTATTCCTGTCCTATTG
TTTTTCTATACCCTTGGAGCTTTTGCTTGCCAAGTTATATTGGGTTCTCGCAGGTCTCTCGGAACC**ATGTCACTGCAAGTCAAAGACCCA**
CGGTCGCGCGGTCGTTCCAAGTCTCCGAGCGGACGCATCCGCGATCGTTCCAAATCTCGTGACAGCCGATTACCCTCTACCGCTCCCGAC
GCTGCGAGGTCATCGGAGAGAAAGTATCTCGCTACCGATGCGAGGGACGACCATCTGCGCTCGAGATCTCGCGGCCCGCGAGATAGCAAC
ACTTCAGTATCTAGTCATCGAAGCCGTTCTCGATACGATGTTTCGGACTCAGCGTCAGAACGGGATGACCGCAAGGACAGATATCTGCGC
AGCGAGCGACGTCGCGACCACTACTACCAATCAGACTCGGGAGAAAGCAAGCTTGCAAAGCGTGATGACAGAAAGGACTATGCGCGGTCG
CCAAATCTTCGACCAGTGACGTATGAGAGTCCCTCGGATGACTCCTACTCAGATACGGACGACGAGGCGCTGGCCTACGGCGATGCGCCT
AGCGATCTGGAGCGTGATTTCTACGGATACAGAAAGCCAGCGCGTGCGTCCTCGCCTCGCGTAGATGGTCCAGTTATGTCCGGGGCGCTC
AATGGTGCTCCTCCAGCCAAGCACGATTCGTCGCGTTCCCGTCATGCTTCAGATGAGGACATTCCCGGTCATCATCCCAGCTATGCCAGA
CCCGGACAGTTCCAGTATGCCATGCCGTCACAGTATGGCCAATTCCAGCCTAGCTATCCTCCGACATCATCTGCACCGCAGTCAAATTGG
GCCCCCATTCCAGAGTGTGAACGACCCGGCTTCGTGCCCCCGTCGTCGCAGGCCGAGACGCAATCCATCCCAGGAGCATTTCCACAACCT
GTGTCGACCTACCCTGAGATGACTGGTGCCCCTGCTGCTCCAGTCTTTCCAATGTCTCAATATGCAAATCTAGGGCAACAGTATCCATAT
GCGGCTCCCAACGGTCGTCCGCTTTCTGTCTCCAATGCTTACACTCCGGCCGTCAGCTCTCCAAATCATCAGCGCGCTTTATCAAGCGAT
GCAAATACGCATCCTGCGTACGCTAATCCTGCATCTTTTCAATACGCCCATGTTGATCCAAATGTGAAATATAGCTCCAAAAATGGCGCT
GTGCCGTATAAGTATTCTGCGACACCCCAGTTCTCCAAACCGAGTGAAGCTCCGACAGCCGAACCCCAACCTACTAGGGCCAAATACTCC
GCAGTTCCACAATTTCAAAAGCCAGCCGCGACAACTCATCAGGAGAATGAACAGCGATTTATCGAGATCACGCCCGGGAGTCGGGCTGCT
GGCCGCCCTGCCAGTCATTCTGTCTCCTCTGCGAACAACCTTTCAGTCGCGGGGACTGATCCAACTCTTCGACCTGCAAGTCCCATGCTA
GAGCCTTATAAGGGCACGTATCAATCTATATCTCCAATGCCGTCGCCTATTGTGATTCCCTCCAAGTTCGACGACGACATCTCTGACTTT
GAGCCACTTGACGGAAGCTCAGATGCGGACGGGAGGAGGAAACACTCGCGAAACAAGTCCAAGGATGAAAGGGATCGCAAAGAGCTGAAG
CCCGAACGCGCTAAACGCGACAGCAGCCGGGTTCGTCACGGGCGACACGAATCGCAAGACCAAACAGTGGTATTGATCTCACCTAGCAGT
GTCCGGAAGAAGGTCACATTCTATGACCCCGGTCCGGATGCGATTACCATGCAGGAGGCTTTGTCGCACACCATCAACATTGATACCAAG
GCTCTCATTCGTGTGCTGCCTCACCTATCCAGCGAGGAAATGCTAGACTTGCGCAAGGAATACAAGAGCCATGTGAAGCTACAGGGCAAA
GGCATCAATATGGCCAAACATATCCGGCTGAAGCTGGGCAACAGTGCGTTTGGGAAAGTATGCTACGCAACCGCCCTGGGCCGCTGGGAG
TCCGAAGCGTACTGGGCCAATTGCTACTACCAGTCTAGCACCTCGCGGCGGGAACTCCTCATCGAGTCTCTAATGGGTCGCAGCAACAGC
GACATCCGCGAGATCAAGCAATGTTTCCGTGACTCACGGTACTTGGACAGCCTGGAAAAGTGCATGAAGGCAGAACTCAAGGCGGACAAG
TTTCGAGTCGCGGTGCTGCTTGCCCTCGAGGAAACACGGCAGAGCGAGCGTGAGCCCATCGATCCGAATCTGGTCCAGCGCGACGTTCGG
GACTTGTATGGGGCCCTGATGTCACGGCACGGGGGCGAGACCGCAATGATCTACATTATTGTACGACGGAGCGACGCACATTTGCGCGAG
GTGCTCCGCGCTTATGAAAAGATCTACAACCAAAACTTTGCTCGTGCAATGATTGCCAAGTCGCAGAATCTTGTGGTACGTATCTCTACT
TTTAGTTCGCCAAGTATCCCTCAAGATCTTCTAACTTGGTTGATGACAG**GGTGAAACACTGGCGCATATTCTCAACGGGGCCATCAACCG**
ACCCATGCGCGATGCTCTGCTGTTGCACCAAGCCCTCCGTGAGAGTCGCTCGGGCAGAGAGCGATCCGAACTTCTAATCTCCCGGTTGGT
CCGTTTGCACTGGGAGCCGCGGCACCTGGAACAGGTTAAGAACGAGTATCGTCGGCGCTATGGGGAACGGCTTGAGGAGGCGATTGCCGA
GGAGATCCTGCCGTCGCCCGGAGGAAGCGAATGGGGCGAGTTCTGTATCGAGTTGGCGCGCAGCTCGAAAACTATTACTGGACGCGGGTG
ACGTCTCTTATCATGCGTAACGATTTTGGGTTATGAGTTATGCGTTATGAGTTGCTTGGTCTTACGACACGGGAATGGTGTTATATTTTC
TTGGCTTTGATTGATATTGGATCAGGTAAAATAAACATATAAATCTTTTCAGGGGTGGATGCATGCTTGACCGGGGTACACCAGTTACTT
ACTGAAAACATCCCTAAAAGCGAAGACTAGTTAACTAACTGGCCCCTAGCTGCTTACTCACATCAAGCATTAAGCGTTAGTTAACTAGGA
GGTCGCTTCCCAAAGCCCATAGGGGATACCGAAGGCGAAATGTCGATTTTCAAAAATGAAAAGTGCTAGATAGAGTTGATATGTAGTATC
TACAAAGTAAGTGTGCGAGTTAAACAAGCAAATATCACTGATTTAGGAGGAAGAGTCTCACTCTCATCCCAATTGCATAGCGACGGCTTT
CGT

Fig. 1 (*Continued*).

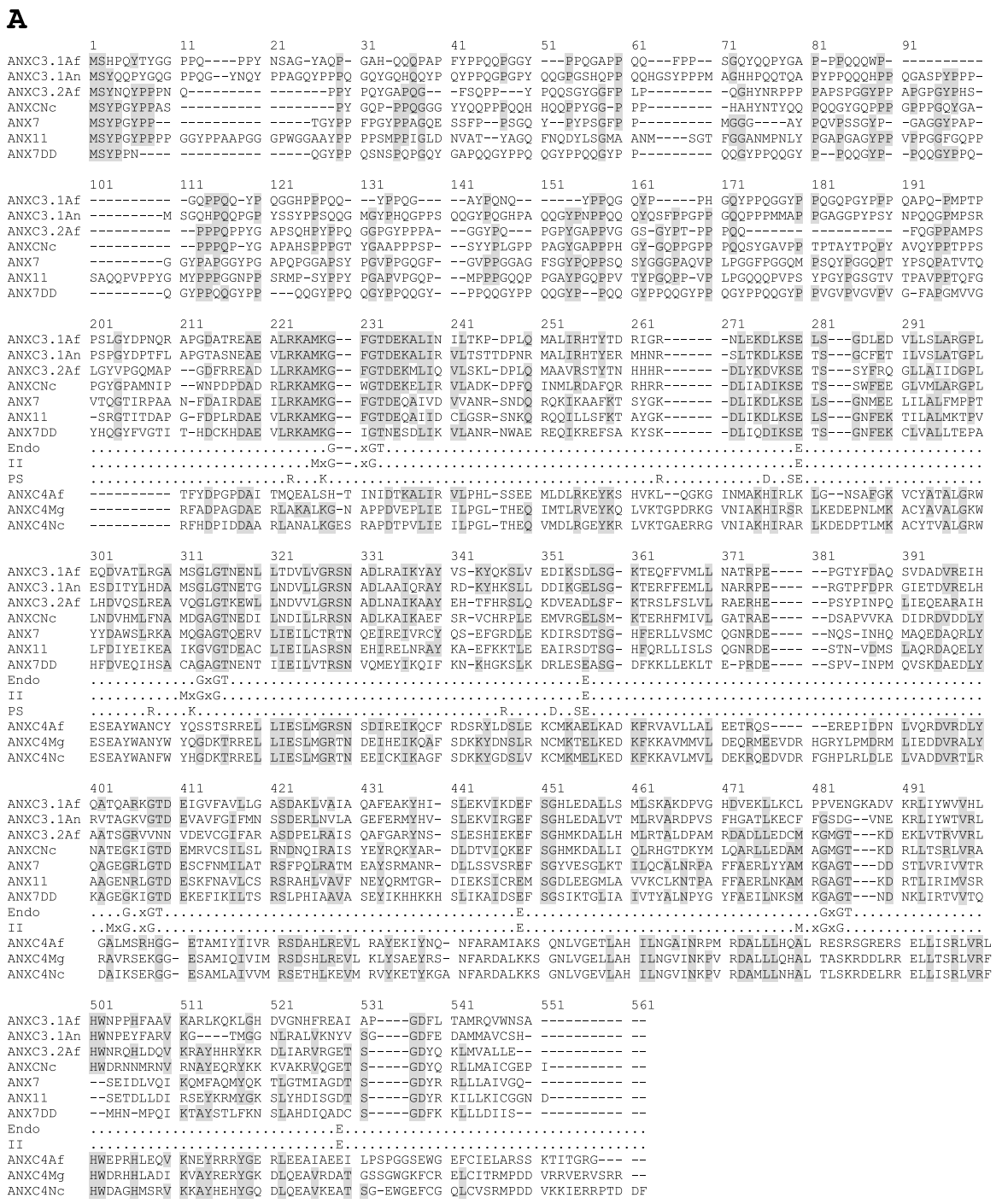

Fig. 2. A: Protein sequences for *A. fumigatus* annexins, shown aligned with other fungal and animal annexins. Sequence name abbreviations are as given in Table 1 with the addition of ANX7DD, annexin 7/ANXC1, *D. discoideum*. 'Endo'; endonexin sequence consensus [29]; 'II', type II Ca^{2+}-binding sites [5,6]; 'PS'; the phosphatidylserine-binding consensus, which is only found in the first two repeats of annexins [6]. In some cases Ca^{2+} binding by particular residues is supplied by backbone carbonyls and so the relationship between sequence consensus and Ca^{2+} binding may be complex [5,6]. B: An alignment of the ANXC4 N-terminal tails. Shading indicates residues identical in at least half of the sequences (A), or in two sequences (B).

tus genome. While cDNA cloning and RACE are dependent on the expression of the message of interest, RT-PCR was also carried out using mRNA extracted from cultures of *A. fumigatus* as the template. This confirmed that all three annexin genes were expressed (Fig. 4).

3.2. ANXC4, a novel class of annexins

The bulk of the ANXC4 gene corresponded to a large open reading frame (the second exon) with the 5′ATG toward the 5′ end of this exon. Since there were a number of stop codons upstream of this ATG and in the first exon, it was concluded that the first exon is not translated. Untranslated exons are unusual, but not without precedent, and examples have been reported previously in fungi [25–27]. The untranslated exon could play a role in the regulation of translation.

Searches of the *N. crassa* and *M. grisea* genome data-

B

```
          1         11        21        31        41        51        61        71        81        91
ANXC4Af    MSLQVKDPRS RGRSKSPSGR IRDRSKSRDS RLPSTAPDAA RSSERKYLAT DARDDHLRSR SRGPRDSNTS VSSHRSRSRY DVSDSASERD DRKDRYLRSE
ANXC4Mg    MSLQVEERRR -GRSRSPGGR DR-------- SRSREPPARV IEASAPYPDD D--------- ---------- ----RDGGVR MYDYED---- ---------P
ANXC4Nc    MSLQVDERSR GGRSRSPADR RREDEDRRPR SHVRTSSANI VEAGVYPPAN DNYRDVDPAL AYEQLANYPP SAGNRGGDYY HSSYHASAPS LGRPRERSPS

          101       111       121       131       141       151       161       171       181       191
ANXC4Af    RRRDHYYQSD SGESKLAKRD DRKDYARSPN LRPVTYESPS DDSYSDTDDE ALAYGDAPSD LERDFYGYRK PARASSPRVD GPVMSG--AL NGAPPAKHDS
ANXC4Mg    RARSYDLPYP APSGRDRD-- ---------- ---------- ---------- ---ARDEPDR FS--YLPQKY AAQSE----- ---------- ----------
ANXC4Nc    RERVRDQSRL RAQSRDRRPD PRESSDWGKN I--------- ---------- ---NRERPGE FAGMYLPPKY ATTVVSADST GSPHGSRDSF HGHGRTPSTS

          201       211       221       231       241       251       261       271       281       291
ANXC4Af    SRSRHASDED IPGHHPSYAR PGQFQYAMPS QYGQFQPSYP PTSSAPQSNW APIPECERPG FVPPSSQAET QSIPGAFPQP VSTYPEMTGA PAAPVFPMSQ
ANXC4Mg    GRSRRDNRDR YD-------- ------DDDL AYGAKEPPPP PSSSSYQY-- ---------- ---------- -------AQP KQYEY----- ----------
ANXC4Nc    GRDRDSSRER QQRKKD---- ----RLEDDL AYGKR--PET PPSPGYHR-- ---------- ---------- -------PRP SSYHAGVG-- --------RP

          301       311       321       331       341       351       361       371       381       391
ANXC4Af    YANLGQQYPY AAPNGRPLSV SNAYTPAVSS PNHQRALSSD ANTHPAYANP ASFQYAHVDP NVKYSSKNGA VPYKYSATPQ FSKPSEAPTA EPQPTRAKYS
ANXC4Mg    -DRADDRLTY SSSRQDSSQL LVAGEPRRRD RSRSPGPSSS SKSYRDDPRG SSASIVTIEP G--------- VSSSASRRSK RDSSPLPPTA R---------
ANXC4Nc    DDKYGEPIGR HSGRTHSPNI SISSAYDLRS DERKGGRIDA PSVVTIEPPH VSGSGYPSSP GRGRHEDDGK LPLKSAMRPG REKSPLPPTN R---------

          401       411       421       431       441       451       461       471       481       491
ANXC4Af    AVPQFQKPAA TTHQENEQRF IEITPGSRAA GRPASHSVSS ANNLSVAGTD PTLRPASPML EPYKGTYQSI SPMPSPIVIP SK----FDDD ISDFEPLD--
ANXC4Mg    -MSSLSVSGG --HHRSG--- ---------- ---------- --SMSLAA-- ---APPSPLL ESYHGTYQSM SPMPSPLMIA SRDPAGHDTR ILDVAPLD--
ANXC4Nc    -MSLLSVNGH NPHMSNG--- ---------- ---------- --NLSVAN-- ---APASPML ESYHGTYQSM SPMPSPLMLP TQ-PGPNGSS YSFVEPLSPL

          501       511       521       531       541       551       561
ANXC4Af    -GSSDADGRR KHSRNKSKDE RDRKELKPER AKRDSSRVRH GRHESQDQTV VLISPSSVRK KV
ANXC4Mg    ---SDDDDRY G------GGD KTKRRA---- ---------- ---------- ---------- --
ANXC4Nc    YGSSSSDDEY HRKGGGSGGG EKKGHARTHS RRA------- ---------- ---------- --
```

Fig. 2 (*Continued*).

bases identified the ANXC4 homologues NCU02139.1 and MG03714.4 which aligned with ANXC4 (Fig. 2A,B). ANXC4 was also identified in the *A. nidulans* and *Fusarium graminearum* genomes (data not shown), and an EST corresponding to ANXC4 was identified in *Mycosphaerella graminicola* (AW180333). ANXC4 is therefore not restricted to *A. fumigatus*.

Comparison of the three ANXC4 proteins with fungal and human annexins indicated many significant differences. The N-terminal tails were 299–553 amino acids long, considerably longer than any other annexin tail (Fig. 2B), with a number of conserved regions. Karabinos et al. [28] recently reported an annexin from the chordate *Ciona intestinalis* (GenBank accession CAE01321) which has an ~400 amino acid intermediate filament sequence fused to the N-terminus. ANXC4 N-terminal tails showed no similarity either to the *Ciona* N-terminus or to intermediate filaments in general as assessed by Blast and PFAM (data not shown); however, these data do suggest

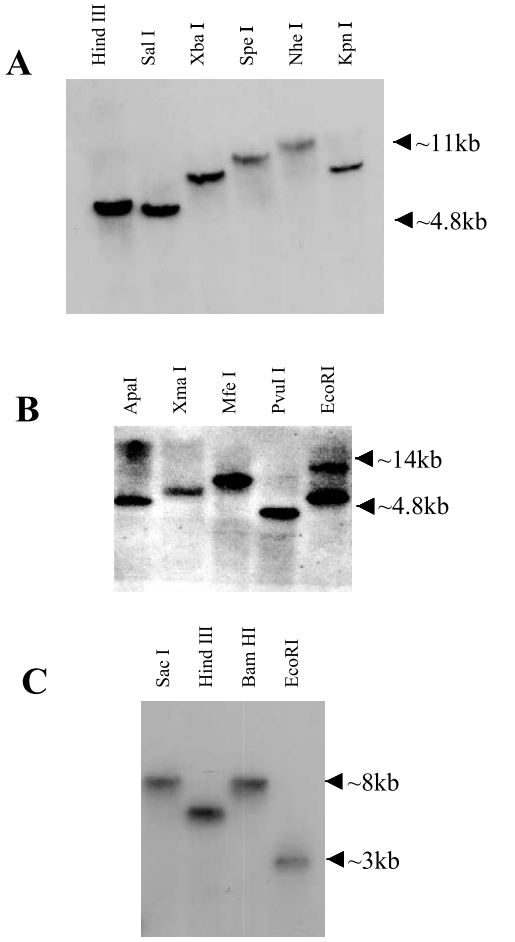

Fig. 3. Southern blots for (A) ANXC3.1, (B) ANXC3.2 and (C) ANXC4. The double band in B is due to *Eco*RI cutting within the ANXB gene, yielding two hybridising fragments.

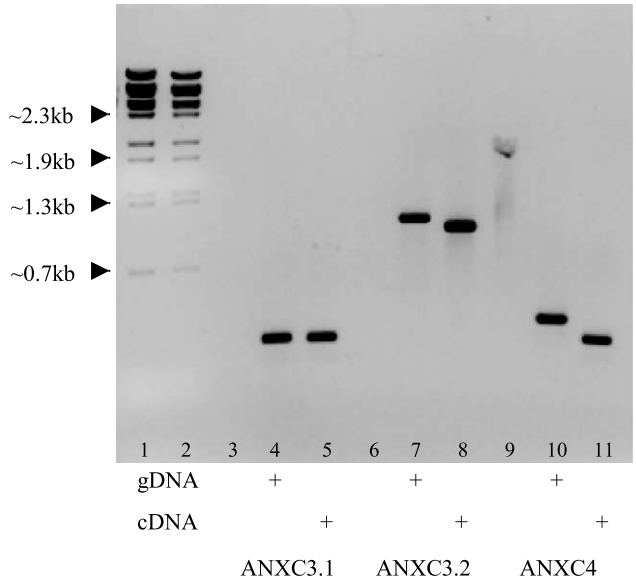

Fig. 4. RT-PCR of mRNA from *A. fumigatus* cultures to demonstrate expression of ANXC3.1, ANXC3.2 and ANXC4. Lanes 3, 6 and 9, no template; lanes 4, 7 and 10, gDNA template; lanes 5, 8 and 11, cDNA template. ANXC3.1, gDNA and cDNA products are both 354 bp; ANXC3.2, gDNA product is 1042 bp, cDNA product is 968 bp; ANXC4, gDNA product is 415 bp, cDNA product is 351 bp. ANXC3.1, primers LS1+LS2; ANXC3.2, primers LS35+LS4; ANXC4, primers LS5+LS6.

Table 1
Percentage identities of fungal annexin core regions against a selection of other annexins

	ANXC3.1Af	ANXC3.1An	ANXC3.2Af	ANXCNc	ANX7	ANX11	ANX7DD	ANXC4Af	ANXC4Mg	ANXC4Nc
ANXC3.1Af	–	54.3	40.7	37.9	34.2	30.6	34.9	27.6	24.9	23.8
ANXC3.1An		–	40.2	38.6	31.3	28.3	34.5	17.8	15.8	14.6
ANXC3.2Af			–	53.7	39.1	31.5	36.5	30.6	28.1	25.6
ANXCNc				–	40.4	35.1	39.3	24.7	24.0	22.1
ANX7					–	53.0	47.9	23.8	22.1	19.9
ANX11						–	40.9	17.3	15.5	13.5
ANX7DD							–	23.7	21.4	20.8
ANXC4Af								–	68.9	67.4
ANXC4Mg									–	80.8
ANXC4Nc										–

ANXC3.1Af, ANXC3.2Af, ANXC4Af: *A. fumigatus* annexins reported here; ANXNc: *N. crassa* annexin [2]; ANXC3.1An: *A. niger* annexin (AY033935_1); ANX7: human annexin 7; ANX11: human annexin 11; ANX7DD: *D. discoideum* annexin 7; ANXC4Nc: *N. crassa* ANXC4, NCU02139.1; ANXC4Mg: *M. grisea* ANXC4, MG03714.4.

that large functional domains can be present as N-terminal tails.

The cores of the ANXC4 proteins aligned with the other fungal and human annexins (Fig. 2A), but only showed 13.5–30.6% identity to them (Table 1). Examination of the consensus for the endonexin sequence, type II Ca^{2+}-binding site and the phosphatidylserine-binding sequence (Fig. 2A) indicated little or no conservation [5,6,29]: the MxGxG sequence was not seen in any ANXC4 repeat and, although co-ordination by these residues is provided by backbone carbonyls, conservation is important for the structure of the Ca^{2+}-binding loop. The downstream 'cap' D/E of the type II binding site, which co-ordinates through the side chain carboxyl, was only seen in repeat II (and in repeat IV of *A. fumigatus* ANXC4). These observations make it unlikely that ANXC4 binds calcium or interacts with membranes in a conventional manner. Recent studies of annexin A9, which also shows poor conservation of type II Ca^{2+}-binding sites, indicated that it only binds Ca^{2+} at concentrations higher than would be seen within the cell [30]. This binding was thought to be via type III sites, which again involve backbone carbonyls, but are not well conserved at the sequence level. Type III sites may be present in ANXC4, but ultimately, Ca^{2+} binding by ANXC4 may have to be assessed experimentally.

In view of the important differences between ANXC4 and other annexins, further analyses were carried out: Blast analysis with the *A. fumigatus* ANXC4 core gave matches to mammalian annexin 8 with *E* values of 1e-11 to 2e-11, as well as to many other annexins, including the *A. niger* annexin AY033935_1 ($E = 9e\text{-}10$). Submission to the FUGUE tertiary structure prediction server (http://www.cryst.bioc.cam.ac.uk/~fugue/prfsearch.html [31]), with the PSI-BLAST option turned off (thereby preventing any sequence matches to annexins before the structural analysis commenced), gave a top match to annexins with a *Z* score of 20.30, the next match scoring only 3.84. Finally, analysis using the PFAM protein fingerprints gave a match to annexins ($E = 6.5e\text{-}4$), with no significant matches to other families. These data confirm that ANXC4 is an annexin.

3.3. ANXC3.1 and ANXC3.2

The core sequences of ANXC3.1 and ANXC3.2 were well conserved compared to other annexins (Fig. 2A,B; Table 1; D. Tuckwell, unpublished observations), showing the endonexin/Ca^{2+}-binding consensus and the phosphatidylserine-binding consensus [6,29]. Both ANXC3.1 and ANXC3.2 had 130–140 amino acid N-terminal tails which showed a high proportion of A, G, P, Q and Y (Fig. 2A) similar to those seen in human and *Dictyostelium* annexin 7 (ANXC1), human annexin 11, and the *A. niger* and *N. crassa* annexins. The calcium-binding protein sorcin interacts with the N-terminal region of human annexin 7 [32], and the related protein ALG-2 interacts with annexin 11 [33]. Searches of the *A. fumigatus* genome database with human sorcin sequence identified a single homologue ($E = 1.6e\text{-}21$). This is therefore a potential binding partner for ANXC3.1 and ANXC3.2.

Previous studies of annexin evolution indicated a single clade (ANXC) made of fungal and mycetozoan annexins [2,23,34]. Searches of the *A. nidulans* genome identified two ANXC3-like annexins, while searches of the *N. crassa* database only identified a single ANXC3-like gene, corresponding to the previously described annexin [2]. Phylogenetic analysis showed that the fungal annexins formed a single clade, with a duplication leading to ANXC3.1 and ANXC3.2 clades (Fig. 5). These data suggest that there should be an as yet unidentified ANXC3.2-like annexin in *A. niger*. The position of the *N. crassa* annexin was not robustly determined but it was clear that this gene did not cluster with either the ANXC3.1 or ANXC3.2 clades. The duplication which led to these two clades in the genus *Aspergillus* therefore probably occurred after the last common ancestor of *Neurospora* and *Aspergillus*.

Searches of EST databases identified ANXC3 annexins in *Coccidioides immitis*, *Colletotrichum trifolii*, *Fusarium graminearum*, *Fusarium sporotrichoides*, *Magnaporthe grisea*, *Pleurotus ostreatus*, *Phytophthora infestans* and *Phytophthora sojae*. No annexin was found in *C. albicans*, *S. cerevisiae* or *S. pombe*, or *E. cuniculi*. Annexins are therefore present in a wide range of fungi, including filamentous fungi, basidiomycetes, and oomycetes (which are not strictly fungi). The absence of annexins from yeasts and microsporidia, coupled with their presence in animals, indicates that annexins were probably lost from one or more ancestral branches of fungi.

In conclusion, we have identified three novel annexin genes in the human pathogenic fungus *A. fumigatus*. ANXC4 is radically different from all other annexins reported to date and defines a novel subset of annexins found in filamentous fungi. The function of ANXC4 remains to be determined, but the clear differences between ANXC4 and other annexins suggest that this gene could be a good target for anti-fungals, although as yet undefined fungal-specific interactions involving ANXC3.1 and ANXC3.2 might also make these good targets for drug design.

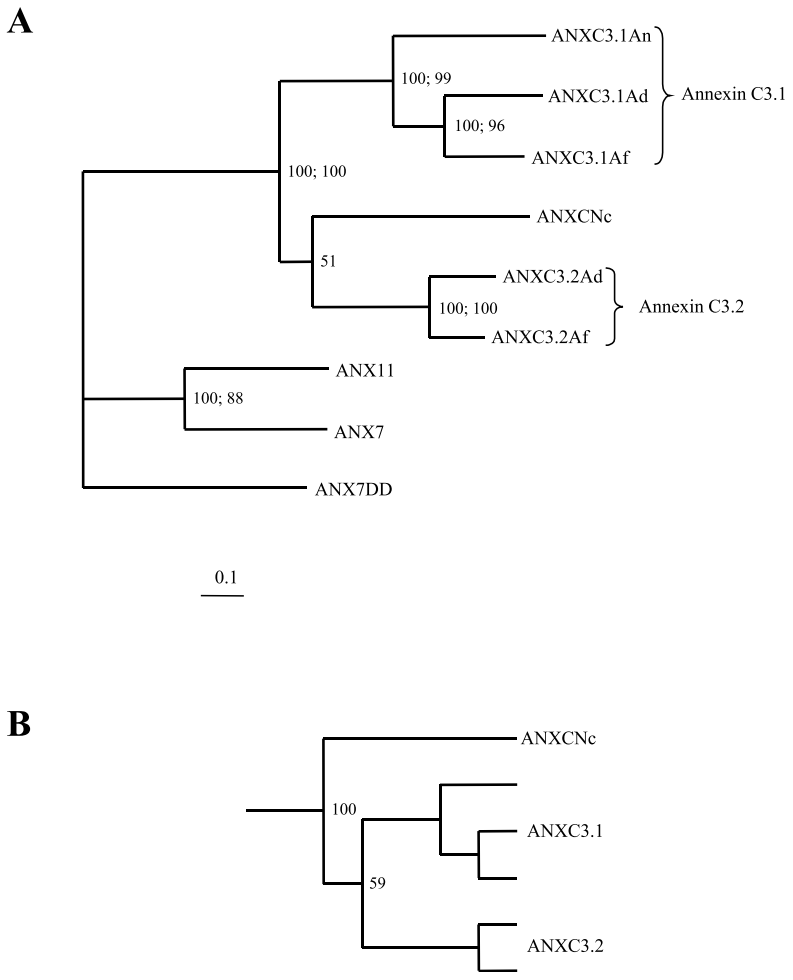

Fig. 5. Phylogenetic tree showing the evolutionary relationships between fungal annexins. A: Distance tree, with bootstrap values from the distance and parsimony analysis of 100 replicates superimposed. At each node the first value is from distance analysis and the second from parsimony. The parsimony tree (1206 steps) had an identical topology with the exception of *N. crassa* annexin (see B) and therefore only the bootstrap value from distance analysis is given for the node at which ANXCNc joins. The scale bar corresponds to the expected number of substitutions per site. B: Detail of the parsimony tree showing the position of the *N. crassa* annexin branch and relevant parsimony bootstrap values. Sequence name abbreviations are as given in Table 1 with the addition of: ANXC3.1Ad, *A. nidulans* annexin 1, contig 1.26; ANXC3.2Ad, *A. nidulans* annexin 2, contig 1.40.

Acknowledgements: We thank Dr R.O. Morgan, University of Oviedo, for guidance on the subject of annexin nomenclature, and Dr G. Robson for helpful discussions. Preliminary *A. fumigatus* sequence data were obtained from The Institute for Genomic Research website at http://www.tigr.org. Sequence data for *C. albicans* were obtained from the Stanford Genome Technology Center website at http://www-sequence.stanford.edu/group/candida. Sequencing of *C. albicans* was accomplished with the support of the NIDR and the Burroughs Wellcome Fund.

References

[1] Gerke, V. and Moss, S.E. (2002) Physiol. Rev. 82, 331–371.
[2] Braun, E.L., Kang, S., Nelson, M.A. and Natvig, D.O. (1998) J. Mol. Evol. 47, 531–543.
[3] Hofmann, A., Proust, J., Dorowski, A., Schantz, R. and Huber, R. (2000) J. Biol. Chem. 275, 8072–8082.
[4] Barton, G.J., Newman, R.H., Freemont, P.S. and Crumpton, M.J. (1991) Eur. J. Biochem. 198, 749–760.
[5] Weng, X., Lueke, H., Song, I.S., Kang, D.S., Kim, S.-H. and Huber, R. (1993) Protein Sci. 2, 448–458.
[6] Montaville, P., Neumann, J.-M., Russo-Marie, F., Ochsenbein, F. and Sanson, A. (2002) J. Biol. Chem. 277, 24684–24693.
[7] Huber, R., Römisch, J. and Pâques, E.P. (1990) EMBO J. 9, 3867–3873.
[8] Swairjo, M.A., Concha, N.O., Kaetzel, M.A., Dedman, J.R. and Seaton, B.A. (1995) Nat. Struct. Biol. 2, 968–974.
[9] Bitto, E., Li, M., Tikhonov, A.M., Schlossman, M.L. and Chow, W. (2000) Biochemistry 39, 13469–13477.
[10] Lambert, O., Cavusoglu, N., Gallay, J., Vincent, M., Rigaud, J.L., Henry, J.P. and Ayala-Sanmartin, J. (2003) J. Biol. Chem. (in press).
[11] Rescher, U., Zobiack, N. and Gerke, V. (2000) J. Cell Sci. 113, 3931–3938.
[12] Raynal, P. and Pollard, H.B. (1994) Biochim. Biophys. Acta 1197, 63–93.
[13] Swairjo, M.A. and Seaton, B.A. (1994) Annu. Rev. Biophys. Struct. 23, 193–213.
[14] Rand, J.H. (1999) New Engl. J. Med. 340, 1035–1036.
[15] Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A. and Barrell, B. (2000) Bioinformatics 16, 944–945.
[16] Galgan, J.E., Calvo, S.E. and Borkovich, K.A. et al. (2003) Nature 422, 859–868.
[17] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D., Smith, J.A., Seidman, J.G. and Struhl, K. (1987) Current Protocols in Molecular Biology, Greene, New York.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

[19] Felsenstein, J. (2002) PHYLIP (Phylogeny Inference Package) version 3.6a3, Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, WA.
[20] Tuckwell, D. (2002) Matrix Biol. 21, 63–66.
[21] Page, R.D.M. (1996) Comput. Appl. Biosci. 12, 357–358.
[22] Morgan, R.O., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Balsara, B.R., Testa, J.R. and Fernandez, M.P. (1999) Genomics 60, 40–49.
[23] Iglesias, J.M., Morgan, R.O., Jenkins, N.A., Copeland, N.G., Gilbert, D.J. and Fernandez, M.P. (2002) Mol. Biol. Evol. 19, 608–618.
[24] Moss, S.E. and Morgan, R.O. (2004) Genome Biol. (in press).
[25] Murasugi, A., Tanaka, S., Komiyama, N., Iwata, N., Kino, K., Tsunoo, H. and Sakuma, S. (1991) J. Biol. Chem. 266, 2486–2493.
[26] Donzelli, B.G., Lorito, M., Scala, F. and Harman, G.E. (2001) Gene 277, 199–208.
[27] Deutsch, M. and Long, M. (1999) Nucleic Acids Res. 27, 3219–3228.
[28] Karabinos, A., Zimek, A. and Weber, K. (2004) Gene 326, 123–129.
[29] Lieman, S. and Huber, R. (1997) Cell. Mol. Life Sci. 53, 516–521.
[30] Goebeler, V., Ruhe, D., Gerke, V. and Rescher, U. (2003) FEBS Lett. 546, 359–364.
[31] Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) J. Mol. Biol. 310, 243–257.
[32] Brownawell, A.M. and Creutz, C.E. (1997) J. Biol. Chem. 272, 22182–22190.
[33] Satoh, H., Shibata, H., Nakano, Y., Kitaura, Y. and Maki, M. (2002) Biochem. Biophys. Res. Commun. 291, 1166–1172.
[34] Morgan, R.O. and Fernandez, M.P. (1997) Cell. Mol. Life Sci. 53, 508–515.